EP 33651 (R)

F2

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 April 2001 (12.04.2001)

PCT

(10) International Publication Number
WO 01/25947 A1

(51) International Patent Classification[7]: G06F 15/16

(21) International Application Number: PCT/US00/27419

(22) International Filing Date: 4 October 2000 (04.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/157,632 4 October 1999 (04.10.1999) US

(71) Applicants and
(72) Inventors: **MEIDAR, Liad, Y.** [US/US]; 2440 Western Avenue, #804, Seattle, WA 98121 (US). **MILLIGAN, George, H.** [US/US]; 33 Interlaken Road, Orlando, FL 32804 (US). **DAY, John, C.** [US/US]; 1461 Willard Road N.W., Palm Bay, FL 32901 (US). **HANCOCK, Monte, F.** [US/US]; 406 Dartmouth Avenue, Melbourne, FL 32901 (US). **MCCORMICK, Rodney, L.** [US/US]; 1208 Walden Pond Court, Rockledge, FL 32955 (US). **FRETHEIM, Erik** [US/US]; 3811 170th Avenue, S.E., Bellevue, WA 98008 (US).

(74) Agents: **HEIMBECHER, Reed, R.** et al.; Dorsey & Whitney LLP, Suite 4400, 370 Seventeenth Street, Denver, CO 80202-5644 (US).

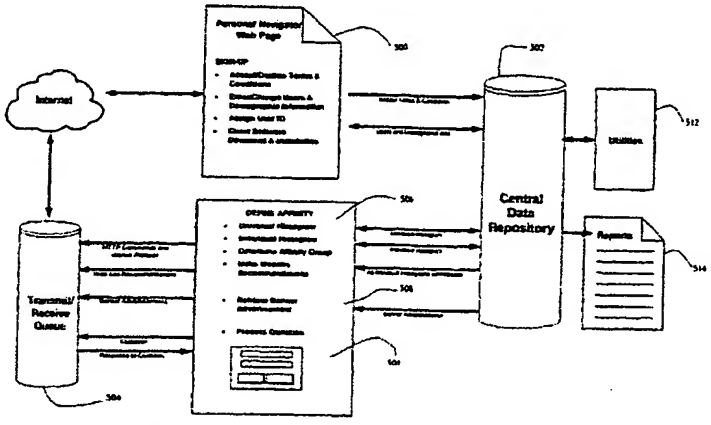
(81) Designated States *(national)*: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD OF DYNAMICALLY RECOMMENDING WEB SITES AND ANSWERING USER QUERIES BASED UPON AFFINITY GROUPS



(57) Abstract: A personal navigation system is disclosed that makes recommendations of information sources of interest to the user. The system makes its recommendations based on complex pattern matching algorithms that take into account the past navigating behavior (506) of the user and behaviors of other users with similar backgrounds who have demonstrated interests in the same concepts. The personal navigation system combines detailed demographic data along with time stamped Web page content to develop a complex waveform (500) representing an individual's usage. As the individual browses the Web, this waveform is matched to other similar waveforms to provide site recommendations. The time stamping feature tracks interests as they change over time. An additional component of the system is a dialogue box (504) that allows the user to query on any subject of interest. The system instantaneously forms an ad-hoc affinity group of other users to which it transmits this query anonymously through e-mail or instant messaging to ask for recommendations from other users in blind or offline open dialogue. Further, by tracking the usage patterns of its users, the personal navigation system can amass valuable marketing data.

WO 01/25947 A1

Method of Dynamically Recommending Web Sites and Answering User Queries Based Upon Affinity Groups

Cross Reference to Related Application

This application claims the benefit of U.S. provisional application Serial No. 60/157,632, filed 4 October 1999, entitled "Method of Dynamically Recommending Web Sites and Answering User Queries Based Upon Affinity Groups and Generating Marketing Intelligence Reports about Consumer Behavior on the Internet Based Upon Accumulated Usage Data." (the "632 application). The '632 application is hereby incorporated by reference as though fully set forth herein.

Field of the Invention

The present invention relates generally to Internet search engine methods and technologies as well as systems and methods for deriving marketing data based upon accumulated Internet usage data for groupings of users.

Background of the Invention

The global computer network known as the Internet has emerged as a mass communications and commerce medium enabling millions of people worldwide to share information, create community among individuals with similar interests, and conduct business electronically. According to International Data Corp (IDC), a market research firm, the number of Internet users will increase from approximately 97 million at the end of 1998 to 320 million by the end of 2002. This exponential growth of the World Wide Web portion of the Internet (the "Web") has made it increasingly difficult for individual users to derive maximum value from the Web. The explosive growth of the Internet is unprecedented in its magnitude, diffusion, and capabilities.

A. Internet Search Engine Systems

Finding information on the Internet is becoming increasingly difficult as the Internet continues to exponentially grow and change. Web sites have proliferated along with the data available on these sites, making it more difficult and time consuming for users to find the

BNSDOCID: <WO_____0125947A1_I_>

information they want. Users are spending a substantial portion of their time searching for the specific information, products, or services they desire. According to IDC, over 100 million Web searches are conducted every day. Furthermore, once an Internet user locates a desired site or sites, the user often finds it difficult to navigate such sites. As Web technology has improved many Web sites have become more complex by adding new features. According to Forrester Research, 1.5 million new Web pages are added to the Internet every day.

Search engines struggle with the dilemma of covering a larger portion of the Internet and providing a large amount of data with poor relevancy versus covering a smaller portion of the Internet and providing a smaller amount of data with greater relevancy. Neither method, though, provides the best of both worlds. Affinity groups and other categorization methods are static or are dependent on human cataloguing. A catalog of some of the presently available searching systems follows. These systems can be grouped according to the foundations of their methodologies as recommendation systems, author-controlled systems, and editor-controlled systems.

1. Recommendation Systems

Recommendation systems attempt to provide the user with additional information about sites visited and related sites of potential interest. For example, Alexa Internet™ is a free Web navigation service that works with the user's browser and accompanies the user when surfing the Web, providing useful information about the sites currently being viewed and suggesting related sites. To use this service, a user downloads and installs software from Alexa™. In March 1999, Alexa™ surpassed 2 million downloads of its software, had 150 advertising partners, and receives 130 million impressions monthly. The company had $370,000 in annual revenues when acquired by Amazon.com™ in April 1999 for $250 million. Alexa™'s technology has also been integrated into both Netscape Navigator® 4.5 and Microsoft Internet Explorer® 4.0 and later versions.

BNSDOCID: <WO_____0125947A1_I_>

eTour™ is another recommendation system that automatically directs people to a different Website, matched to their unique interest, each time they connect to the Internet and reward users for its use. This personalized, ever-changing service has signed up over 325,000 people. eTour™'s recommendation system is powered entirely by humans and does not involve any sophisticated technology. All recommendations are for Websites that have paid eTour™ to recommend their site to a targeted audience.

Direct Hit Technologies is a provider of popularity-based, recommendation system, search engine products. This engine ranks results according to which sites millions of Internet searchers have found useful. Direct Hit™ helps searchers broaden or narrow their search by displaying additional search topics that others have found helpful for similar searches. This technology also has personalization features that enable users to receive search results tailored to their gender, age, or geographic region. Direct Hit™'s revenues are derived from licensing their technology to clients such as Lycos™, Hotbot™, MSN™, ICQ™, Looksmart™, among others.

2. Author-Controlled Systems

The "author-controlled" search engines such as Inktomi™, Alta Vista™, Excite™, Infoseek™, and Lycos™ work by comparing the words in the search request with the words in the millions of available Web documents. While these engines are capable of automatically locating a large amount of information, they have not proven reliable at reducing this large body of information down to a manageable set of documents relevant to the search request.

Moreover, author-controlled engines essentially empower the authors of documents to control their own ranking by the words that they choose to put into their Web site documents. Because a ranking on the highly trafficked search engines means more traffic to the site owner, site owners have an incentive to get the highest possible placement for their site,

BNSDOCID: <WO_____0125947A1_I_>

regardless of its quality with respect to other sites or its relevancy to any particular search request. This conflict of interest has caused many site authors to strive for the highest placement possible by tricking the search engines.

For example, Inktomi™'s search technology powers many Internet search engines, which include regional or global Internet searching, retrieval systems for large text archives and powerful online search support for publisher archives. This search application is optimized to handle the combination of massive data and larger user bases, without requiring the use of expensive multiprocessor supercomputers. Inktomi™ scientists recently developed a new technology (Concept Induction Technology), which uses supercomputing techniques to model human conceptual classification of content and projects this intelligence across millions of documents.

Alta Vista™ is a deep search spider that indexes all the pages within a Website. The company uses a ranking algorithm to determine the order in which matching documents are returned on the results page. Each document gets a grade based on how many of the search terms it contains, where the search terms are in the document, and how close to each other the search terms are. Alta Vista™ also uses site popularity to help boost the ranking of a Website.

Excite™ is different from other navigation tools because it uses "concept searching," understanding language to the point that it uses synonyms. Excite™'s spider summarizes each page using sentences, which express its dominant concept. Pages are then reviewed and automatically rated. Keywords are assigned to a page according to what the spider deems is the page's theme. Excite™ assigns a "confidence rating" according to how closely the queried words match what it considers to be the theme of a site.

A variation of these author-controlled search engines is emerging which factors in link-popularity. These engines, such as Google™ and Clever™, rank documents higher

BNSDOCID: <WO 0125947A1_I_>

based on the number of links pointing to the document and the text around those links. These methods are essentially adaptations of the time-honored tradition of rating articles based on the number of citations and operate to empower other authors to determine the ranking of sites. The usefulness of such systems is limited, however, because a large number of sites lack a meaningful number of referring links, and other authors often link to sites for a variety of unrelated reasons. Not unexpectedly, the Web promotion industry encourages site owners to link together to increase their rankings. Google uses a complicated mathematical analysis, calculated on more than a billion hyperlinks on the Web, to return high-quality results. Google prioritizes search results by how many other pages it has found that link to the page in question, and thus it judges how important or authoritative the rest of the people making up the Internet consider that page to be.

3. Editor-Controlled Systems

To counter the effects of allowing authors to control the rank of their own sites, search sites such as Yahoo™, LookSmart™ and About.com™ use an editorially created directory of sites. These "editor-controlled" sites each employs a staff of editors to manually select and catalog Web sites. Even the aforementioned author-controlled search engines have added similar editorially created directories above their traditional search result listings. By slowly adding sites to the index, these companies build an index of Web documents that are each carefully reviewed before addition to the index.

The amount of labor needed for such a task, however, is quite high. While the quality of this body of data tends to be much higher, this expensive and labor-intensive process is incapable of keeping up with the constant growth and change in the Web. Reports indicate that these directories have managed to catalog only a small fraction of the Web so far and that the review process for submission of a site can take up to several months.

BNSDOCID: <WO_____0125947A1_I_>

Moreover, most users actually navigate this much smaller body of editorially created data with a conventional search engine. Because these directories are usually organized alphabetically within generic categories, a conventional search engine is needed to search across categories to find requested information. So while the editorial staff reduces the author's ability to wrongfully influence a site's ranking, the conventional search engine's ranking algorithms still provide a large number of irrelevant results. The problems of both conventional search technologies and editorially created directories only get worse for e-commerce searching. The listing delays and the high cost of labor prohibit the use of an editor-controlled model for organizing individual product pages.

B. Internet Marketing Systems

The Internet enables advertisers to target advertising and marketing campaigns utilizing sophisticated databases of information about the users of various sites and to directly generate revenues from these users through online transactions. As a result, the Internet has become a compelling means to advertise and market products and services. IDC estimates that global e-commerce revenue is expected to increase from approximately $32 billion in 1998 to more than $425 billion in 2002. According to Meyers Group, online advertising expenditures will reach $32 billion by 2005 up from $2 billion in 1999.

Existing media research firms include Media Metrix, Nielsen Media Research/NetRatings, Jupiter Communications, and All Advantage.com. These companies offer a wide array of products and services that include marketing information and identifying Internet opportunities and trends which could be useful in estimating market demographics and demand curves.

For example, Media Metrix provides Internet audience measurement products and services to leading Internet advertisers, advertising agencies, Internet properties, technology companies, and financial institutions. Media Metrix collects this data by measuring Internet

BNSDOCID: <WO_____0125947A1_I_>

usage from a representative sample, or panel, of personal computers with their proprietary metering system, which is contained in a software application installed on a panelist's personal computer. The meter monitors all communications between the computer's operating system and the software applications and hardware that the operating system controls and monitors.

Nielsen Media Research and NetRatings formed a strategic alliance joining their separate Internet audience measurement initiatives to create a new service to provide the media industry with information on how people are using the Internet. This partnership provides Internet advertisers, marketers, site publishers, media planners, and Web professionals with comprehensive information about Web user interaction with Web sites and ad banners. The company's data collection technology captures detailed Web usage information from a randomly-recruited, statistically-representable group of Web users and compiles behavioral data with in-depth demographic and lifestyle profile information.

C. Internet Application Toolboxes

In addition to the methodologies of the search engine and marketing data systems described above are application "toolboxes," plug-ins, and other supporting technologies, which provide added functionality to the search engine and marketing systems. Other companies with technologies of interest along these lines include Net Perceptions™, Artificial Life™, Ask Jeeves™, Inquizit™ Technologies, IBM®, Third Voice™, and Hypernix™.

Net Perceptions™ is a developer and supplier of real-time recommendation technology that enables Internet retailers to market to customers on a one-to-one basis. Real-time technology predicts an individual's preferences and makes specific recommendations accordingly. The technology does this by learning about each individual's preferences through observing relative behavior, recalling past behavior, and asking the individual to rate

BNSDOCID: <WO_____0125947A1_I_>

a number of relevant items. The technology then pools this information with knowledge gained from a community of other individuals who share similar tastes and interests. This technology integrates collaborative filtering, neural networks, fuzzy logic, and genetic algorithms.

Artificial Life™ is a provider of intelligent software bots for Internet applications. Artificial Life™ develops intelligent software bots that can multi-task across the enterprise with the effective use of natural language. The bots are designed for user convenience and the automation of business-related Internet and intranet tasks including Web navigation, direct marketing, user profiling, information gathering, messaging, knowledge management, sales response, and call center automation. Artificial Life™ is also developing products for data mining, Web page analysis, statistical analysis, and direct marketing to support the functionality of Artificial Life™'s suite of intelligent software products.

Ask Jeeves™ is a provider of natural-language question answering services on the Internet for consumers and companies. The Ask Jeeves question answering services allow users to ask a question in plain English and receive a response pointing the user to relevant Internet destinations that provide the answers.

Inquizit™ Technologies' patented software linguistically interprets the meaning and concepts of plain English. This technology analyzes sentence structure, grammar, word meanings, and content. It incorporates a dictionary that contains most of the common English words and all their different meanings in addition to using a dictionary of concepts and incorporates common sense knowledge.

IBM's Intelligent Miner for Text™ (IMT) is a software development tool kit. This product offers the ability to extract patterns from text, organize documents by subject, and research for documents that match a given topic. IMT has text analysis tools and an advanced search engine enhanced with mining functionality and capabilities to visualize

BNSDOCID: <WO_____0125947A1_I_>

results. IBM is also currently working on a search technology research project named "Clever."

Third Voice™ enables online discussion forums for private, group, or public interaction. This service allows users to freely and openly express ideas at points of references anywhere in a Web page using a free browser companion.

Hypernix™ is an Israeli company dedicated to developing innovative communications tools. With Gooey™, its recently released freeware product, Hypernix™ introduced the concept of Dynamic Roving Communities (DRC), which allows Internet users the opportunity to interact and communicate anywhere, and anytime on the Web. More than 50,000 users have downloaded this hybrid of Web surfing and chat technology since June 14, 1999.

D. Limitations of the Prior Art

While the growth of the Internet has drawn users at an unprecedented pace, the volume of online information has made it increasingly difficult for users to navigate the Internet effectively. As mentioned above, Forrester Research estimates that 1.5 million new Web pages are added to the Internet every day. To take full advantage of the Web, users must be able to successfully navigate a network of dispersed Web sites, which are generally not connected in a logical fashion.

Users currently rely on Internet search engines or directories of Web sites and Web pages to locate information and find sites of interest. Search engines typically require consumers to construct keyword or complex search strings that often result in hundreds or thousands of matches. As directories become larger, they require users to move through large and complex hierarchies of information. As the Internet grows, users of conventional search and directory products are finding that locating the information they need is increasingly difficult.

BNSDOCID: <WO_____0125947A1_I_>

The problem with most search engines is that they return an overwhelmingly large number of Web sites for each search request. For example, one of the major search engines returns over 163,720 possible Webs sites for the search request "Boston Car Dealerships." Obviously, it is impossible for a person to look at all of these sites, and many people spend a significant amount of time just trying to find one or two Web sites which relate to their search request. As a result, people have become frustrated using conventional search engines.

Another major shortcoming of search engines is their insufficient coverage of the Internet. Three popular directories, Lycos™, Excite™ and Yahoo™ respectively cover only 2.5%, 5.6%, and 7.4% of the Internet, and the Internet continues to grow faster than these companies can index sites. In fact, the most comprehensive search service, Northern Light only covers 16% of the entire Internet. So these search engines may miss sites that are both popular and relevant to a particular search.

In order to fulfill the promise of the Internet, access to information, products, and services of interest to the user must become simpler and quicker. Until navigation on the Internet improves, users will become increasingly frustrated with their online experiences. What is needed is a more direct and personal means of interacting with the Internet that will improve the user's experience and enhance companies' returns from Internet strategies. This will, in turn, make the Internet more valuable to users and companies alike.

Although search engines are the most common search method, according to eStats 53% of users still rely on recommendations from friends and relatives as one of the most common navigations methods. What is needed then is a service which complements that of search engines; a service that points out to the user the most popular sites visited by people within the user's demographic pool, and which match the areas of interest to the user. In addition, it would be desirable for a user to be able to instantly communicate with an appropriate "affinity group" to seek answers to questions or information needs. Affinity

BNSDOCID: <WO_____0125947A1_I_>

groups are developed by the personal navigation system based on common interests and/or backgrounds as defined by usage patterns and demographics.

The number of people who are "surfing" the Internet is growing exponentially. As mentioned above, IDC predicts that the number of Internet users will increase to 320 million by the end of 2002. As the Internet grows to become the central nervous system of global commerce, companies marketing or conducting business over the Internet will require increasingly higher level of commercial intelligence. What is needed is market intelligence reporting that will provide much greater insight into consumer behavior then that which exists today.

Summary of the Invention

The invention disclosed herein comprises a personal navigation system that uses advanced artificial intelligence technology that transforms the way users presently navigate, communicate, and find relevant information on the Internet. The personal navigation system, as a permission-based consumer and business tool, also changes the way companies reach their prospective customers, as well as provides some of the most revealing information available about consumer behavior on the Internet.

The personal navigation system provides a nonobtrusive window adjacent to the user's main browser (e.g., Internet Explorer® or Netscape Navigator®), through which the personal navigation system makes recommendations of sites that would be of interest to the user. The personal navigation system makes its recommendations based on complex pattern matching algorithms that take into account the past navigating behavior of the user and behaviors of others with similar backgrounds who have demonstrated interests in the same concepts to create groupings based upon affinity between users. The personal navigation system combines detailed demographic data along with time stamped Web page content to develop a histogram that is translated into a complex waveform representing a user's usage.

BNSDOCID: <WO_____0125947A1_I_>

The complex waveform of a user is the DNA of that user's web behavior. It describes the user's interests at their most basic level, i.e., it is a collection of "key words" or "atomic phrases" that represent "the meaning" of the Web pages they have visited. As the user browses the Web, the user's waveform is matched to other similar waveforms to provide information source recommendations. The time stamping feature tracks interests as they change over time. Unlike other search engines, directories, or other navigation tools, the personal navigation system provides a personalized and active approach to recommending sites, which others have found to be useful or interesting.

An additional component of the personal navigation system is a dialogue box. The personal navigation system dialogue box allows the user to query on any subject of interest. The personal navigation system instantaneously forms an ad-hoc affinity group to which it transmits this query anonymously through e-mail to ask for recommendations from other users who have demonstrated an interest in the topic in question. Recipients can, if they choose, respond anonymously and, if both parties choose, can engage in blind or offline open dialogue. The personal navigation system dialogue box will allow users to communicate with their peers anonymously around the world to get recommendations on sites which other users have found useful or to answer other questions.

In order to use the personal navigation system, users will, upon registering, download personal navigation system tracking software from a central data server. This tracking software will collect and transmit back to the central data server anonymous data on the Internet usage of each user. Since the personal navigation system tracks behavior of others most like the user, the automatic personal navigation system recommendations and dialogue box answers are apt to be more relevant and accurate than any prior art search or affinity group.

BNSDOCID: <WO_____0125947A1_I_>

Further, by tracking the usage patterns of its users, the personal navigation system can amass valuable marketing data. The personal navigation system uses sophisticated data mining techniques to reveal significant details on consumer behavior on the Internet. The personal navigation system can use this marketing data to generate marketing intelligence reports. The personal navigation system is able to provide answers to such questions as "When are mothers most likely to buy books?" as well as competitive information, such as "How successful is my competitor's banner or TV ad?" Since the personal navigation system can create very specific affinity groups (e.g., a group of women in their forties who live in New York City with a certain income level who are interested in French cooking and hiking), true targeting is possible. In another aspect of the present invention, these conclusions will be summarized in the marketing intelligence reports, which are sold to e-commerce and online marketing companies trying to determine the best way to target potential clients. Marketing intelligence reports can answer key subtle questions regarding specific consumer behavior, such as "Which sites are most visited by Ivy-League bankers?" or "How effective was my competitor's TV ad on the Super Bowl for influencing consumer Web behavior?"

The foregoing and other aspects, features, details, utilities, and advantages of the invention will be apparent from reading the following description and claims, and from reviewing the accompanying drawings.

Brief Description of the Drawings

Figure 1 is a schematic representation of a preferred network for implementing the personal navigation system of the present invention over the Internet.

Figures 2A and 2B depict a flow process for registering and creating a waveform a user within the personal navigation system of the present invention.

Figure 2C is a schematic representation of various matching methodologies performed by the personal navigation system of the present invention.

BNSDOCID: <WO_____0125947A1_I_>

Figures 3A and 3B depict a flow process for implementing an anonymous dialogue between users of the personal navigation system of the present invention.

Figure 3C is a schematic representation of an anonymous dialogue between users of the personal navigation system of the present invention.

Figure 4 is depicts a flow process followed by the personal navigation system of the present invention for mining user data and creating marketing reports.

Figure 5 is schematic diagram of the various functional components of the server end of the personal navigation system of the present invention.

Figure 6 is a representation of a universal histogram according to the personal navigation system of the present invention.

Figure 7 is a combined representation of a user waveform and histogram according to the personal navigation system of the present invention.

Figure 8 is a schematic diagram of the various functional components of the user end of the personal navigation system of the present invention.

Detailed Description of the Preferred Embodiments of the Invention

A method and system of dynamically recommending Web sites and answering user queries based upon affinity groups is now described in detail and with particular reference to preferred embodiments. The most preferred embodiment comprises a personal navigator system for the Internet. The detailed description further discusses methods for generating marketing intelligence reports about consumer behavior on the Internet based upon accumulated usage data within the context of the personal navigator system.

Figure 1 depicts the overall architecture of the personal navigation system of the present invention, which follows two distinct technical paradigms. A portion of the personal navigation system operates in a thin-client environment through an industry standard browser on the user's computer 100 accessing the personal navigation system Web site 110.

3NSDOCID: <WO_____0125947A1_I_>

preferably over the Internet 120. The remaining portion executes on a combination of both the user's computer 100 (client) and the personal navigation system server 110. Both the client and server execute independently with a point-to-point communication link being established on-demand for exchange of information, when appropriate. Other possible network configurations, for example distributed networks, local area networks, wide area networks, and the like, are well known in the art and may be alternately used to implement the processes of the invention. Also, while the personal navigator system is described in particular reference to Web pages and similar information sources on the Internet, its system and method are similarly applicable to use on other communication networks where information sources are available and sought by users. Such other communication networks may include intranets, private and public networks, ATM networks, telephony networks, and broadcast, cable, and satellite television. For example, numerous other information sources are accessible over the Internet and transferred via Internet protocol packets. Other information sources are available via telephony networks. A further example is the in-band and out-of-band information transmitted in television broadcasts, most notably in vertical or horizontal blanking intervals.

Figures 2A and 2B depict the steps taken by a user to register with and begin using the personal navigation system. First a user accesses a Web site hosting the personal navigation system 200, for example, the Personal Navigator™ system soon to be available from Personal Navigator, Inc. The user then completes a questionnaire 202 including requests for name, address, e-mail address, gender, birthday, occupation, income level, marital status, number of children, and college attended. The user further checks off boxes indicating areas of general interest to the user. Next the user reads terms of use and privacy statements 204, opts whether to receive targeted e-mail information and advertisements 206, and clicks a button, indicating acceptance of the conditions 208.

BNSDOCID: <WO_____0125947A1_I_>

Once the user accepts the conditions, the personal navigation system creates an anonymous user ID 210 and automatically downloads the personal navigation system tracking software onto the user's computer 212. As the user navigates through the Internet, the tracking software captures data on the user's behavior 214, including, but not limited to, the following: selected text from Web sites visited (preferably including nouns, but not adjectives); time of use data (date, day of week and time of day) and duration of viewing for each page viewed; frequency of hits by all users on each site viewed; and the path taken to reach each page viewed (e.g., by collecting HTTP commands). While the user is connected to the Internet, the personal navigation system completes a periodic (daily or as often as the user connects to the Internet) "quiet" upload of the user's usage data 216 collected by the tracking software on the user's desktop. The user is generally unaware of this event. The personal navigation system then employs pattern-matching algorithms 218 to determine the following: concepts of interest to the user 220; sites matching concepts of interest to the user viewed most commonly by the user's affinity group 222 (developed by the personal navigation system based on common interests and/or backgrounds as defined by usage patterns and demographics); sites most commonly viewed by the user's affinity group 224; and sites on the Internet containing concepts of interest to the user 226. See Figure 2C.

Once the user's initial demographic information, interests, and affinities are known and/or calculated, the personal navigation system can instantiate its recommendation functions. Figures 3A and 3B depict the steps of the ensuing process. When the user next connects to the Internet and opens a browser 300 (e.g., Internet Explorer® or Netscape Navigator®), the personal navigation system window opens along-side the user's browser window 302. As the user navigates through the Internet, the personal navigation system window displays a list of recommended of sites to visit 304, and links thereto, based on the user's current navigation matched against sites with similar concepts historically visited most

BNSDOCID: <WO_____0125947A1_I_>

commonly by members of the user's affinity group, or presently being visited by members of the user's affinity group. The personal navigation system window may also display products of potential interest to the user based upon concepts of interest to the user and products viewed by members of the user's affinity group.

The personal navigation system further opens and displays a dialogue box 306 in which the user can enter any question, comment, or other message of interest 308. The message is broadcasted anonymously to the personal navigation system dialogue box of all personal navigation system users who have indicated or demonstrated an interest in the concept(s) contained in the message 310. When other users open their personal navigation system browser, a message flag is present to indicate that they are in receipt of an anonymous message from another personal navigation system user 312. Users in receipt of a broadcasted message choose whether to respond anonymously to the sender of the original message or whether to ignore it 314. When a user responds to an original message, a direct, two-way dialogue thread can be created between two anonymous users with a similar interest 316 if both are simultaneously connected to the network.

For example, in Figure 3C a first user interested in Cajun cooking may enter a query in the dialogue box seeking good Cajun recipes 318. A second user of the personal navigation system in the first user's affinity group, simultaneously connected to the system via the Internet, receives the first user's message broadcast by the system, and enters a response 320. The personal navigation system, recognizing that both the first and second users are presently, simultaneously connected to the network, provides a real-time dialogue thread between the users to support ongoing, anonymous communications 322.

A further aspect of the personal navigation system is its ability to "mine" data on anonymous consumer behavior over the Internet to create marketing intelligence reports that assist e-commerce companies to define marketing strategies. Marketing intelligence reports

BNSDOCID: <WO_____0125947A1_I_>

may be used, for example, to optimize text, colors, and placement of on-line ads; understand customer behavior in navigating the Internet; and send targeted e-mail (for those users who have opted to accept email advertising).

Referencing Figure 4, the personal navigation system collects significant data 400 on each user, including demographic information such as age, gender, zip code, income level, marital status, number of children, and education; Web sites and pages viewed; time (date, day of week, and time of day) and duration of each page viewed; the content of each page viewed (e.g., nouns and proper nouns identified on the page), including advertisements but not graphics; the path taken to reach each page viewed; and the frequency of page views. The personal navigation system then applies advanced data mining techniques to analyze the data captured from each personal navigation system user 402. This allows the personal navigation system to make insightful conclusions about consumer behavior on the Internet 404. These conclusions are summarized in various customized marketing intelligence reports 406. It is estimated that about 50,000 users provide the personal navigation system with a "critical mass." This critical mass translates to very pertinent Web site recommendations and marketing intelligence reports.

Marketing intelligence reports prepared by the personal navigation system may take the form of "syndicated reports" and "customized reports." Syndicated reports comprise general information on Web usage, which may be offered through subscription and distributed periodically (weekly or monthly). These syndicated reports may include, for example, information about: unique visitors to most popular Web sites; time (date, day of week and time of day) and average duration of usage; average unique Web pages visited per day and per month; demographic compositions of Web users; purchasing tendencies of Web users; and other behavioral data, including mass consumer trends.

BNSDOCID: <WO_____0125947A1_I_>

The customized reports offer more in depth and revealing analysis of the data collected by the personal navigation system. These reports are prepared and sold individually by request of the customer. These custom reports offer answers to very specific and difficult questions about consumer behavior on the Internet. Examples of such questions topically arranged, which the personal navigation system can answer, are the following:

Industry & Sectors	Who is my biggest competitor? What is the size of my market? Which players excel in which areas?
Customer Retention	To where do my customers attrite? What marketing programs have been successful for my competition?
Consumer Behavior	When is a certain consumer most likely to buy a certain product? How do different demographic groups shop on the Internet?
Site Content	What types of users are attracted to sites with certain content? What content within specific sites attracted the most attention from users?
Site Interaction	What sites have significant overlapping user bases? To where should my site be linked to attract new clients?

Businesses can use the syndicated and customized reports for many purposes such as planning, buying and selling advertising, developing e-commerce strategies, understanding consumer behavior, gaining competitive market intelligence, and analyzing investment decisions.

The technology comprising the personal navigation system consists generally of modern, state-of-the-art software development tools, software languages, communication protocols, and commercial off-the-shelf Web server and browser technology augmented with three key technologies. While the modern state-of-the-art components will certainly play an

BNSDOCID: <WO_____0125947A1_I_>

important role in the implementation of the personal navigation system, it is the three key technologies that create an advantage over similar systems and methods.

The first of these technologies is a natural language parsing tool that enables the identification and capture of key text contained within a Web page that collectively represents the "meaning" of the Web page. The second technology uses a combination of several very multi-dimensional clustering algorithms that enable comparison of the multiple users Web usage (i.e., comparison of the meaning of each Web page), computation of a measure of their similarity, and derivation of affinity groups that represent collections of users with similar interests. By then comparing the list of Web sites visited by members of the affinity group, Web sites visited by other members of the affinity group can be recommended to each member.

The third technology is actually a collection of several algorithms and techniques that perform pattern recognition, data analysis, clustering, classification, inductive learning, and other intelligent information processing tasks. This collection of algorithms enables discovery of very sophisticated and complex relationships within Web usage that can provide valuable market intelligence reports to e-commerce and on-line marketing companies. The following is a brief description of each of these key technologies.

A. Natural Language Parsing Tool

A natural language parsing tool is a software scripting language that performs many sophisticated functions, such as pattern matching and manipulating text and relational objects. It has powerful pattern-matching and string manipulation functions for "drilling down" into text files and decomposing them into relational objects. It also has built-in text formatting functions for "rolling up" relational objects, string concatenation, and producing text reports.

BNSDOCID: <WO_____0125947A1_I_>

In the most preferred embodiment, that natural language parsing tool is called RML, available from Computer Science Innovations, Inc. (CSI) of Melbourne, Florida. Originally, RML stood for "Report Markup Language" because it was designed as a markup language to extract, manipulate, and format text strings into textual reports. Later, it was enhanced and used to analyze textual rules in a knowledge based expert system, breaking down each rule into its component parts, creating a common grammar, and subsequently restructuring each rule to fit within the established grammar. Thereby, it became "Rule Making Language." As RML continued to be enhanced, mature, and be used in a variety of different problem domains, it became clear that RML's real strength lies in its ability to recognize and manipulate relations, thus RML became "Relational Methods Language." Most recently, RML was used for machine learning and data mining projects. For example, a system designed completely in RML was used to identify patterns in medical ledger information and subsequently convert medical charge descriptions into coded numbers for a large data warehouse company.

B. Multi-Dimensional Clustering Algorithms

The human brain is the most sophisticated pattern-matching machine known to mankind. Viewing Web pages, understanding their meaning, and recognizing similarity or non-similarity between Web pages is a task that can be easily and quickly accomplished by the human mind. Yet, comprehending the "meaning" of a Web page and subsequently deriving a measure of similarity is an extremely difficult task to perform automatically in software. In the most preferred embodiment, the clustering algorithms comprise part of an Advisor Toolkit, which is also available from CSI.

CSI's Advisor Toolkit contains a variety of mathematical algorithms that possess the power to distinguish similarities and differences within the content of Web pages. The most significant of these is a neural paradigm, which, through unsupervised clustering, maps Web

BNSDOCID: <WO_____0125947A1_I_>

page content into multi-dimensional feature space and subsequently computes a measurement of "nearness" based on a variety of mathematical metrics. In effect, these algorithms can build a profile that accurately represents a user's Web behavior. This capability enables recognition of similar Web usage and similar interests among different users and thus, the establishment of affinity groups and recommendations of potential Web sites of interest.

C. Information Processing Algorithms

In the most preferred embodiment, these routines collectively comprise the Advisor Toolkit mentioned above. The Advisor Toolkit is a collection of advanced software routines that perform pattern recognition, data analysis, clustering, classification, inductive learning, and other intelligent information processing tasks. These routines enable exploitation of the repository of Web usage information collected by the personal navigation system. Combined into a hybrid system, these powerful routines can perform detailed and in-depth analysis of Web usage data for knowledge discovery and identification of hidden patterns and correlations within the usage patterns. From these analyses, market intelligence reports, customer behavior reports, and predictions of behavior based on the historical behavior and profiles of individual users, and communities of users, can be produced.

Before proceeding to the detailed functional descriptions, definitions of several technical terms are required.

Atomic Phrase: An atomic phrase is defined as a string of text that if subdivided, destroys its semantic associations. For example, "tennis" and "chief operating officer" are both atomic because "tennis" cannot be subdivided and subdivision of "chief operating officer" produces phrases that generally do not preserve its connotation. "French

3NSDOCID: <WO_____0125947A1_I_>

cooking" is not atomic, because many connotations survive subdivision.

Semantic Association: The semantic association of two phrases is defined as one or more words that, when used alone in separate Web searches, return URL lists having many items in common. When used together, they return an URL list largely disjointed from the common URL lists created when the words were used alone.

The software residing on the personal navigation system server performs several primary functions: 1) sign-up, 2) communications, 3) definition of affinity groups, 4) processing of banner advertisements, 5) messaging (i.e., anonymous e-mail), 6) utility routines, and 7) reports. Each of these functions may be implemented as separate software modules and interact through either internal message passing or via event queuing. Figure 5 provides a pictorial view of how these functions interact.

The sign-up module 500 enables a user to subscribe to the personal navigation system service. The sign-up module is invoked by connecting to the personal navigation system Web server using an Internet browser (e.g., Internet Explorer® or Netscape Navigator®). After selecting the "Sign-Up" option, the personal navigation system "Terms and Conditions" is displayed within the browser window. Acceptance or non-acceptance of the Terms and Conditions is indicated by the subscriber pointing and clicking on either an "Accept" or "Decline" button. The "Decline" button displays an appropriate completion message and exits the sign-up module. The "Accept" button triggers several activities that prepare for initiation of the personal navigation system on the user's computer.

Upon acceptance of the "Terms and Conditions," a formatted data entry panel is displayed within the browser into which the subscriber enters two pieces of information: the

BNSDOCID: <WO_____0125947A1_I_>

textual name of each user within the household that will be using the computer and demographic information describing each person. Completion of entry of the information is indicated by the subscriber by, for example, pointing and clicking on a "Next" button. All textual names and demographic information are stored in a data repository 502 on the personal navigation system server. The list of textual user names is also stored on the client computer. Storage on the client computer facilitates easy selection and change of the current user of the system without connecting to the personal navigation system Web server. As the number of users grows to a significant number, the amount of data in the data repository 502 will also grow to a substantial size. The design and structure of the data repository 502 is preferably upwardly scalable to accommodate timely storage of data in the data repository 502, as well as able to provide expedient search and retrieval functions to quickly identify affinity group members.

Each user's computer is assigned an identifier that uniquely and anonymously identifies the computer. To obtain a truly unique identifier, the identifier is preferably a concatenation of several pieces of information including one of the textual names, a date/time stamp, with time expressed to the millisecond, and a one-up numerical suffix. The one-up numerical suffix is established and controlled by the personal navigation system server, thus allowing each number to be accessed only a single time and guaranteeing a unique identifier even contemplating the slim chance of multiple identifiers being created with the same textual name during the same millisecond. This unique identifier is appended to the textual name of each individual user of the computer, thus uniquely identifying each user. Acceptance of the "Terms and Conditions," the textual names, and associated demographic information describing each individual user is stored in the central data repository 502 residing at the personal navigation system Web server, also with a concatenation of the unique identifier.

BNSDOCID: <WO____0125947A1_I_>

Download of the client software and user identifiers is again accomplished via standard browser functionality using hypertext transfer protocol (HTTP). A message is displayed, with a progress bar, indicating the download is being performed. Upon completion of the download, installation of the client software is automatically initiated by the browser. Automatic installation of downloaded files is currently supported by both Internet Explorer® version 5 and the current release of Netscape Navigator®. Third party products for automatic installation after download are also available that install as plug-ins to both browsers. The installation script does not require user interventions with the exception of specifying the drive and directory onto which the personal navigation system client software is to be installed. A default selection is suggested.

The communications module is responsible for sending and receiving all information from and to the Web server. Three types of information are transmitted: 1) Web site recommendations derived from an affinity group; 2) banner advertisements; and 3) responses to questions posed to affinity group members. Three types of information are received: 1) HTTP commands and atomic phrases; 2) requests for banner advertisements; and 3) questions to be posed to affinity group members. All communications involving transmission of information are initiated by the server and are accomplished via a point-to-point (PTP) connection. All communications involving receipt of information are again accomplished via a PTP connection, initiated, however, by the client computer. All PTP connections, regardless of the originator, are established and remain open for only the duration of the transmission.

The communications module operates based upon receive and transmit queues 504. The type of information in the queues is identifiable, preferably by the file type. All received information is queued for processing by the other modules. The communication module periodically examines its "transmit queue" and if a file or files are present, transmits the

BNSDOCID: <WO_____0125947A1_I_>

file(s) to the appropriate client computers. The communications module has responsibility for re-try of transmissions when communications fail or transfers are interrupted. When the transfer is complete, the transferred files are deleted from the server.

The definition of an affinity group is the hub of the personal navigation system. The define affinity group module 506 examines all atomic phrases and, through a series of processing steps, enables comparison of the Web usage of different users and computation of a measure of similarity between users. Users determined to be similar are thus members of an affinity group. Affinity groups are ad hoc dynamic associations; they vary with time and the nature of the comparison being performed.

Affinity group definition is accomplished by the creation of a histogram that captures the universe of atomic phrases. The histogram is itself a list of all the atomic phrases captured from all the Web pages that have been visited by all users of the personal navigation system. When files containing atomic phrases are received at the personal navigation system server, each atomic phrase is compared to the atomic phrases residing within the existing universe. If an atomic phrase is not present in the histogram, it is added to the universe. If the atomic phrase is already present, no action will be taken. Only one occurrence of each atomic phrase is present in the histogram. Thus, the universal histogram can be described as a sequential list of every atomic phrase encountered to date. Note that an atomic phrase is comprised of any combination of the characters within the ASCII character set. Thus, the universal histogram is completely insensitive to language and context. A histogram representing a small world might be similar to that shown in Figure 6.

A similar process is followed to create each individual user's histogram that describes their Web usage activity. Each user's histogram is a concatenation of their demographic information and their list of atomic phrases. All demographic information is represented in the histogram as numeric values. This requires the set of demographic information to be a

BNSDOCID: <WO_____0125947A1_I_>

closed set of information. Demographic information such as gender, age, marital status, and state of residence is captured and consistently described by numeric values. These values are simply transcribed into text when displayed. Other information, such as special interests and hobbies is selected from a closed list of information, thus also allowing numeric representation of this information.

The list of atomic phrases is actually a list of pointers that point to the location of the atomic phrase within the universal histogram. Again, there is only a single occurrence of each unique atomic phrase in the user's histogram. Each pointer is accompanied by other relevant information including the following: 1) a date/time stamp reflecting when each atomic phrase was encountered; 2) a count which reflects the number of times the atomic phrase has been encountered; and 3) the URL of the Web page from which the atomic phrase was extracted. By considering both the date/time stamp and the counts as a measure of the weight of importance of the atomic phrase, i.e., the more recent encounters being of more importance and the higher count of encounters being of more importance, the set of atomic phrases with the highest weighted value indicate a user's high level of interest in that topic. Figure 7 depicts a sample histogram for an individual user.

Note that there are several other important subtle aspects of an individual's histogram that are of great significance. Each histogram can be viewed as a waveform with the amplitude of the wave at any point in the histogram being the numerically coded demographic information or the weighted combination of count and recency. These waveforms, as points in Euclidean space, provide the mathematical basis for formation of affinity groups using a neural technology called autoclustering. By proper selection of the weight assigned to the recency of encounter with respect to the current date, a natural decay of the significance of each atomic phrase occurs. This natural decay very accurately depicts an individual user's change of interests over time and thus, a change of affinity groups as

BNSDOCID: <WO_____0125947A1_I_>

personal interests change. As a general matter, demographic information should remain relatively static while the atomic phrases and their frequencies representing Web usage are very dynamic.

The content of the set of atomic phrases extracted from a Web page constitutes the entire "meaning" of the Web page. No external interpretation of each page is required. Atomic phrases must be accurately and sufficiently recognized to successfully capture the "meaning" of the Web page. In the preferred embodiment, meaning is extracted from a web page, or other information source using "cognitive engineering." Cognitive engineering is the technology associated with the design and implementation of computerized applications that emulate intelligent human behaviors, such as decision-making, plan development, and problem solving.

Cognitive engineering technology applies cognitive engineering tools according to a "Cognitive Engineering Methodology (CEM). CEM is an objective, systematic methodology for developing systems having embedded intelligence (e.g., neural nets, expert systems, and regression models). This methodology consists of seven steps: 1) problem evaluation and analysis; 2) feature extraction and enhancement; 3) sampling; 4) data analysis and modification; 5) model design and development; 6) model evaluation; and 7) system implementation, testing, and validation. Steps four through six above are repeated, as required. A "spiral development methodology" based on a rapid prototyping approach is often appropriate. The purpose of CEM is to provide a consistent framework within which the two components of data mining can be carried out: "knowledge discovery" and "predictive modeling." Knowledge discovery occurs in steps 1 and 2 above. Predictive modeling occurs in steps 4 - 7 above.

Knowledge discovery is the isolation and characterization of actionable information from data. It consists of problem evaluation and analysis, feature selection, and feature

BNSDOCID: <WO_____0125947A1_I_>

enhancement. The first step, as indicated, is problem evaluation and analysis. Problem evaluation begins with interviews of domain experts, and the collection of raw domain data from accessible repositories. A problem description is written by the system developer, and OLAP tools are used to assess the available data sources for quality and information content. Some techniques used to support the data analysis phase are first, conventional and statistical techniques such as: population modeling by statistical moments (e.g., means and standard deviations); correlation (e.g., testing data for dependence or independence); chi-square analysis (e.g., test hypotheses vis-à-vis the statistical character of the data); simple visualization (e.g., histograms, scatterplots, graphs, and charts); time-series analysis (e.g., control charts, linear predictive modeling). A second group of techniques are online analytical processing (OLAP) techniques such as: stratification and segmentation ("slicing and dicing" data); roll-ups (summarizing data in various ways to seek explanatory patterns); drill-down (expanding summarized data to seek explanatory patterns); complex queries and browsing (to uncover complex relationships); scientific visualization; and sophisticated feature plots and ad hoc data views. A third group of techniques can be described as high-end analysis such as: autoclustering (determining natural patterns in the data); rule induction (generating predictive/explanatory rules from data); and link analysis (discovering significant connections among data).

The second step in the knowledge discovery process is feature extraction and enhancement. Feature extraction is the process whereby data are characterized for processing by pattern recognition and exploitation tools and techniques. Some techniques used during the feature extraction process are elementary conventional methods such as: counts, ratios, differences, and quotients; integral transforms; Fourier transforms (e.g., windowed fast Fourier transforms); wavelets (multi-resolution decomposition); and general kernel filters (spectral, spatial, and temporal). Other techniques can be classified as quantization and

BNSDOCID: <WO_____0125947A1_I_>

coding, such as: MAX quantization, histogram equalization, and view-through-feature coding. More techniques include semantic feature extraction such as tokenization (parse tree) and bag-of-words, as well as regression features such as model coefficients (e.g., slope of least-squares line).

Feature enhancement is the process of transforming and coding data in such a way as to make the information it contains more accessible for automatic exploitation by predictive models. Some techniques used for feature enhancement are: Bayesian analysis (How well will a linear classifier do on this problem?); feature registration and normalization (e.g., z-scoring); excision, replication, and synthesis (e.g., class collisions and population imbalance); feature correlation and salience(Do the features provide independent information?); principal component analysis (PCA) (e.g., Karhunen-Loeve); independent component analysis (ICA); filling in missing data fields (e.g., intra-vector regression); and rule induction (RML, LCR, and BAM (e.g., BOLTZ routines)).

Predictive modeling is the automated exploitation of actionable information based upon the results of the knowledge discovery phase. It consists of sampling, data analysis and modification, model design and development, model evaluation, and system implementation, testing, and validation.

Sampling, which is the third step of the CEM, is the first step to fall under the protective modeling phase. Four statistically representative random samples are created from the data set conditioned in steps 1 and 2. These random samples are the calibration set, the training set, the validation set, and the hold-back set. The calibration set is used to estimate statistical and informational theoretic parameters for the problem (e.g., ranges, minimum values, maximum values, variances, scores, counts, and entropies). The training set is used to construct regression models from adaptive algorithms (e.g., neural networks). The validation set is used to perform "blind tests" to determine the ability of the predictive model to

BNSDOCID: <WO_____0125947A1_I_>

generalize. The hold-back set is retained in a blind store, and used for final validation of the completed model.

The fourth step is data analysis and modification. Once features have been extracted and the development sets are created by sampling, another round of analysis and data transformation is conducted. The same techniques and tools used in step 1 above are applied to the refined data sets to create enhanced sets that will serve as the basis for the final predictive model.

The fifth step is model design and development. Based upon the complexity of the prediction problem, a predictive modeling paradigm (e.g., neural network, expert system, or black-box regression) is chosen. This selection is based upon the analysis conducted in earlier stages of the process, and is a matter of engineering judgment. Typically, less complex problems in well-understood domains are addressed using conventional techniques (e.g., logistic regression and expert systems), while hard problems in poorly understood domains are addressed using advanced adaptive algorithms (e.g., neural networks). Numerous tools for the automatic construction of predictive models may be used, such as model based ("white-box") techniques, decision trees (GINI and two-ing), and non-model based ("black-box") techniques. Model-based techniques can include the following: hard analytic models (e.g., ad hoc mathematical) and knowledge-based expert systems (e.g., forward and backward chaining). Non-model based techniques can include the following: neural networks (e.g., backpropagation and reinforcement learning); multi-layer perceptrons; Hopfield nets (e.g., Boltzmann machines); feature maps (e.g., Kohonen); adaptive resonance theory (ART) machines; regression machines; restricted coulomb energy (RCE) machines; radial basis functions (RBF); adaptive logic networks (ALN); hybrid systems (e.g., "bagging"); and CART-like systems (e.g., INDUCE and SPLITS)

BNSDOCID: <WO_____0125947A1_I_>

The sixth step involves model evaluation. To evaluate a predictive model, it is applied to the validation set (a "blind test"), and scored for performance (typically for classification accuracy or optimization of some objective function). If the results are "good" (a subjective judgment), a further validation may be performed by combining the calibration and validation sets, and using n-fold cross validation. Some other techniques used for model validation are sensitivity analysis and application to "use cases."

The seventh step in the CEM process involves model implementation, testing, and validation. When a deliverable level of performance is achieved, a delivery version is constructed. This version is tested on the hold-back set. In the context of the personal navigation system of the present invention, the CEM process is used capture the "meaning" of the Web page. In the most preferred embodiment the CEM tools in the CSI Advisor Toolkit are used to extract the pertinent content from the information source, a Web page in the preferred embodiment, and create the atomic phrases.

As a general matter, Web pages tend to change with relative frequency. Thus, the atomic phrases that constitute the meaning of a particular Web page are dynamic and will change over time. A natural decay of the importance of atomic phrases that no longer appear in a particular Web page will occur due to the measure of importance assigned to the recency of encounter. If the Web page changes so that the atomic phrases in the new set are different from those in the old, the site is not the "same site" for the purposes of the personal navigation system, even though the URL is the same. The counts and recency of the new atomic phrases will indicate importance and form the basis for the affinity group.

An affinity group is a nonpersistent group of users whose histograms are similar at a particular moment in time. Affinity groups will continuously change and any individual user will likely be a member of multiple affinity groups, each of which represents a different collection of interests. The obvious question is how similar is "similar enough" for multiple

BNSDOCID: <WO_____0125947A1_I_>

users to reside in the same affinity group. The fundamental technique used to determine similarities and create affinity groups is called autoclustering. In the most preferred embodiment, an autoclustering technique called Weighted Pair-Group Centroid is used by the personal navigation system to determine the similarity of users for placement in affinity groups. A detailed description of this and other autoclustering techniques can be found at http://www.statsoftinc.com/textbook.stcluan.html, which is hereby incorporated by reference as though fully set forth herein.

The Weighted Pair-Group Centroid clustering algorithm assumes each entry in a user's histogram is a point in a feature space where the numeric value in each histogram is a coordinate in the feature space. In this way, using N values from a user's histogram represents that user as a point in an N-dimensional feature space. This allows the mathematics of Euclidean N-space to be applied to the analysis of a user's demographics and histories of visitations to information sources, e.g., Web pages.

Affinity groups are created by matching users in the N-dimensional feature space through clustering. This is a four-step process. First, histogram features are selected for use in creating an affinity group for a user or class of users. This is done by designating interests or demographic attributes that are of interest for the match. The default for user affinity grouping is to select the J attributes of the user having the highest histogram counts.

Second, the number of users, M, needed to form the desired affinity group is determined. This will usually default to a statistically relevant value (e.g., 0.01% of the total population), or a similarity threshold (e.g., a maximum distance beyond which individuals cannot be regarded as similar for the purposes of the clustering). Third, each of the N-dimensional points (histogram values) in the entire population to be searched is assigned an initial weight of one. Fourth, the two N-dimensional points (histogram values) having the smallest distance between them are found. These two points are replaced by a single point

BNSDOCID: <WO_____0125947A1_I_>

located at their weighted mean, having a weight which is the sum of the weights of the original two points. This process is continued until exactly M points remain. More generally, any computable function can be used as a "similarity measure". If the Euclidean distance is used, conventional autoclustering is based on nearness results.

In an alternative embodiment, the personal navigation system architecture allows the fourth step above to be replaced with the following, more general, affinity clustering step. In the alternative embodiment, a computable function (the objective function) is applied to the original user for which the affinity group is being formed. The affinity group is called the value so obtained, V. This same computable function is then applied to each of the M N-dimensional points (histogram values); call these values W_i. The list $|W_i - V|$ is sorted in ascending order, and the first (i.e., smallest) K elements are selected. These are the K points having function values most like the function value of the original user, and their members are selected as the affinity group

The alternate embodiment of the fourth step allows complete generality in the formation of affinity groups to support clustering for "arbitrary groups. Arbitrary groups are affinity groups that are, for example, most alike in interest, most alike in web access schedules (apart from interest), most demographically similar, most similar in likelihood of some future behavior (e.g., purchasing, fraud, default), or most similar in an abstract sense.

To derive information source recommendations, the Web sites visited by the K closest neighbors in multi-dimensional space of all members of the affinity group will be compared to the web sites visited by the user. K is chosen in such a way as to give a reasonable number of recommendations; a typical value might be K=10. Web sites, or other information sources, not visited by the original defining user are garnered from the K affinity members, and prepared for transmission to the user's computer. Recommendations are preferably made based upon the popularity of the Web site, with more popular Web sites being recommended

BNSDOCID: <WO_____0125947A1_I_>

first. Popularity is defined as a combination of the number of times the Web site is accessed, the recency of the access, and the dwell time at the site.

This same process is used when questions are posed by the user in the dialogue box embodiment. The atomic phrases within the question are concatenated with the user's demographic information forming a mini-histogram representing only the very limited universe of the question. The mini-histogram is autoclustered as above to identify histograms (and so, other users) that are similar to the mini-histogram. The web sites visited by these other users may be recommended to the poser of the question.

Banner advertisements are associated with specific affinity groups, and therefore, groups of atomic phrases. These associations are established manually by members of the personal navigation system staff by determining atomic phrases that represent the "meaning" of the banner advertisement. An individual's histogram is examined to determine regions of interest. Banner advertisements associated with regions of high weight are selected for transmission to the client computer. This function is implemented in the personal navigation system by a banner advertisement module 508.

The process question module 510 implements the dialogue box features of the personal navigation system, preferably through traditional electronic mail capabilities. When questions are received, the atomic phrases within the question are concatenated with the submitter's demographic information forming a mini-histogram representing only the very limited universe of the question. The mini-histogram is autoclustered and compared to the individual histograms that are nearby in multi-dimensional space. Histograms that are similar to the mini-histogram, i.e., histograms that contain high weights for the same, limited set of atomic phrases, are deemed eligible for receipt of the question. The Web sites visited by these individuals are preferably recommended to the poser of the question. An electronic mail message is programmatically formed and sent to each recipient.

BNSDOCID: <WO_____0125947A1_I_>

When responses are received, incoming electronic mail messages are simply forwarded on to the submitter of the question. If the submitter is not on-line, the server stores any responses until such time that the submitter is on-line and the message can be sent. As the personal navigation system is preferably required to retain the anonymity of users, the process question module 510 translates user identities into the appropriate e-mail address. Thus, a recipient of emails will only have access to the alias user identity information, not the actual e-mail address of the sender.

The personal navigation system also preferably allows question submitters and responders to initiate a two-way dialog thread through which they can communicate directly. This is accomplished using Instant Messaging (IM) technology, originated by America Online. IM has recently become an industry standard means for establishing point-to-point communications between two on-line users. IM technology is available through several commercial sources.

Several utility routines provided by the utilities module 512 are preferred to effectively support the personal navigation system. These include but are not limited to the following: displaying a list of users and demographic information; displaying affinity groups; and displaying banner advertisements and attaching banner advertisements to an affinity group.

The report module 514 provides the ability to mine consumer's Internet usage behavior to create market intelligence and other reports, as deemed appropriate. The primary tool used to identify usage trends is a collection of advanced software routines that perform pattern recognition, data analysis, clustering, classification, inductive learning and other intelligent information processing tasks. In the most preferred embodiment, these routines collectively comprise CSI's Advisor Toolkit. These routines enable exploitation of the repository of Web usage information collected by the personal navigation system. Combined

NSDOCID: <WO_____0125947A1_I_>

into a hybrid system, these powerful routines perform detailed and in-depth analysis of Web usage data for knowledge discovery and identification of hidden patterns and correlations within the usage patterns. From these analyses, market intelligence reports, customer behavior reports, and predictions of behavior based on the historical behavior and profiles of individual users, and communities of users, can be produced.

The software downloaded to the user's computer performs a variety of functional tasks. These may be combined in a single software module or may be separated into individual modules. In either event, these modules interact through either internal message passing or via event queuing. Figure 8 provides a pictorial view of how these functions interact.

The capture module 800 is responsible for intercept, parsing, and storage of Internet usage. The capture module executes as a plug-in to the computer's default browser. The capture module activates based upon two specific events posted by the browser: 1) the sending of an HTTP command; and 2) the receipt of an HTML tag. When a send HTTP command event is posted, the HTTP command is captured. When receipt of an HTML tag event is posted, the HTML tag is captured and parsed for atomic phrases. The set of captured atomic phrases constitutes the entire "meaning" of the Web page.

Preferably, the atomic phrases are captured primarily from the HTML header, anchor, center, title, header tags, and potentially other tagged fields such as a table. Experimentation and evaluation of Web page content has shown that information contained within these tagged fields are normally rich in meaning. Nontext objects (such as images and sounds) will not be captured. However, expansion to include parsing of the actual text comprising the Web page may be necessary, especially with the growth of Extensible Markup Language (XML), which allows definition of localized tags.

BNSDOCID: <WO_____0125947A1_I_>

In the preferred embodiment, parsing of the HTML tags and extraction of atomic phrases are accomplished via a combination of the RML tool and well-established techniques within the natural language parsing domain. RML performs the basic parsing task of separating the narrative text into sentences, phrases, and individual words. Each word then goes through several distinct processing steps. The residual after the processing steps represents the atomic phrases, and thus the meaning of the Web page.

The processing steps in sequential order are as follows. Case is checked to see if the word is comprised of all upper case letters, which would indicate an acronym. All acronyms will be considered atomic phrases because they have been found to hold special significance. A dictionary lookup is next performed. Entries in the dictionary are common words (such as "a," "is," and "the") that hold no meaning. Dictionary matches do not become atomic phrases. Punctuation is deleted with one exception, a hyphen. Hyphenated words contain semantic meaning that is lost when the words are separated, whereas all other punctuation does not add meaning to the stem word. Stemming is next applied to remove plurals and other variances of words that again do not contribute to meaning. Stemming results in a common form of each word. Lastly, another dictionary lookup is performed to exclude socially unacceptable words. Additional processing steps may be beneficial to further refine the residual list of atomic phrases. This step-by-step approach can easily incorporate additional processing steps that may include rule base processing to accommodate special situations and identification of significant word pair associations.

It also may be beneficial to assign a "weight" to each word that enhances or decreases the word's importance with respect to other atomic phrases within the Web page. Atomic phrases with a higher weight would have a higher representation of the meaning of the page. If weighting of words is desired, it is preferred to count the frequency at which the atomic phrase occurs, both within the individual Web page, and also throughout the universe of

BNSDOCID: <WO_____0125947A1_|_>

atomic phrases. With regard to an individual page, a higher frequency of occurrence would indicate a higher level of importance. With respect to the universe of atomic phrases, the inverse is true. A lower frequency of occurrence within the universe indicates a higher level of importance with regard to an individual Web page.

Both the HTTP commands and atomic phrases are saved in local storage 802 as disk files on the user computer in a specifically named "transmit" subdirectory which acts as queue for all data to be sent to the personal navigation system Web server. All saved information is date and time stamped and tagged with the user identifier.

It is important to note that only the contents of displayed Web pages will be parsed. Technically, the entire Web site could be parsed by following the trail of HTTP commands embedded within the HTML tags that comprise the Web page itself. However, the purpose of the personal navigation system is to capture actual Web usage and therefore, the system limits parsing to only displayed Web pages. Programmatically drilling-down into the depths of a Web site does not reflect actual usage.

The communications module 804 is responsible for sending and receiving all information from and to the user's computer. Three types of information are transmitted: 1) HTTP commands and atomic phrases; 2) requests for banner advertisements; and 3) questions to be posed to affinity group members. Three types of information are received: 1) Web site recommendations derived from an affinity group; 2) banner advertisements; and 3) responses to questions posed to affinity group members. All communications involving transmission of information are initiated by the user's computer and are accomplished via a PTP connection. All communications involving receipt of information are again accomplished via a point-to-point (PTP) connection, initiated, however, by the server. All PTP connections, regardless of the originator, are established and remain open for only the duration of the transmission.

BNSDOCID: <WO_____0125947A1_I_>

The communications module operates based upon receive and transmit queues. The type of information in the queues is identifiable, preferably by the file type. All received information is queued for processing by the other modules. The communication module periodically examines its transmit queue and, if a file or files are present, transmits the file(s) to the personal navigation system server. The communications module has responsibility for re-try of transmissions when communications fail or transfers are interrupted. When the transfer is complete, the transferred files are deleted from the user's computer. Incoming files are queued in a manner that enables recognition by the type of data they represent (i.e., affinity information or banner advertisements) and are subsequently pre-processed and displayed by the personal navigation system graphical user interface.

The graphical user interface (GUI) module 806 provides several functions including displaying Web site recommendations based on an affinity group, acting as a receiver for questions posted to affinity group members, displaying responses to posed questions, selecting the active user, changing user and/or demographic information, and displaying banner advertisements. In the preferred embodiment, the personal navigation system GUI executes as its own window and is completely independent from Web browser operation on the user's computer.

The GUI module recognizes that Web site recommendations have been received. The GUI displays the recommendations and initiates the flashing of a personal navigation system icon indicating new information has been received.

The GUI also provides a text window, the dialogue box, into which short, concise questions can be entered. Questions are then queued for transmission to the personal navigation system server. When responses to questions are received, or when questions are received from other affinity group members, the GUI provides an electronic mail capability that queues received messages and allows users to view the messages at their leisure. Most

BNSDOCID: <WO_____0125947A1_I_>

normally recognized email functions are implemented. In particular, the "Instant Message" capability, mentioned above, is implemented, which allows two personal navigation system users to establish a direct point-to-point link between their computers, thus allowing direct communication between the two subscribers in a completely anonymous manner.

The GUI module recognizes that new banner advertisements have been received. The GUI displays the banner advertisements in a manner viewable by the user. The GUI also provides a means for selecting the current user of the computer. A list of individual users resides on the client computer and is displayed in a manner that provides for easy selection of the desired user.

The GUI further provides a means for adding/changing individual users and demographic information, as well as a means for deleting an individual user. A button on the GUI invokes the default Web browser, connects to the personal navigation system Web site, extracts the appropriate demographic information from the database, and displays the information in a formatted manner enabling changes to the information. This screen is preferably the same screen that is used for initial entry of demographic information at the time of sign-up. An "OK" button designates completion of the change and the new information is saved in the data repository. A "Cancel" button aborts the update and no changes are made.

While the system and method of the present invention have been described as encompassing numerous features, capabilities, architectures, and configurations that are depicted herein in some detail, it should be appreciated that the system and method of the present invention encompass any and all combinations of these and features, capabilities, architectures, and configurations, and is not to be construed as limited to any preferred embodiment. Modifications may be made to the processes, techniques, equipment, and other elements disclosed herein without departing from the scope of the present invention.

BNSDOCID: <WO_____0125947A1_I_>

Claims

We claim:

1. A system for dynamically recommending at least one new information source to a first user, the system comprising

a first user system comprising a processor and a memory;

a plurality of second user systems, each comprising a respective processor and memory;

a server system comprising a processor and a memory; and

a communication network linking the first user system, the plurality of second user systems, and the server system; wherein

the first user system stores in its respective memory both demographic information about the first user and historical information identifying previously-selected information sources of the first user, and each of the plurality of second user systems stores in its respective memory demographic information about, and historical information identifying previously-selected information sources of, respective second users; and wherein

the first user system and the plurality of second users systems each transmits its respective demographic information and historical information to the server system via the communication network, and wherein the server system stores the transmitted demographic information and historical information in the server system memory; and wherein

the server system processor creates respective histograms for the first and second users from the stored respective demographic and historical information, and the server system creates respective waveforms for the first and second users by transforming the respective histograms into the respective waveforms based upon frequencies indicated by the respective histograms; and wherein

BNSDOCID: <WO_____0125947A1_I_>

the server system processor compares the waveform of the first user to the respective waveforms of the plurality of second users; and wherein

the server system processor selects a subset of second users from the plurality of second users, wherein the waveforms of the subset of second users indicate an affinity among the subset of second users and the first user; and wherein the server system processor recommends at least one new information source to the first user based upon the previously-selected information sources of the subset of second users.

2. The system of claim 1, wherein the historical information further comprises a specific time of selection for each previously-selected information source.

3. The system of claim 1, wherein the historical information further comprises a duration of selection for each previously-selected information source.

4. The system of claim 1, wherein the historical information further comprises a frequency of selection for each previously-selected information source.

5. The system of claim 1, wherein the historical information further comprises a sequence of selection for each previously-selected information source.

6. The system of claim 1, wherein the historical information further comprises at least one content indicator for each previously-selected information source.

7. The system of claim 6, wherein the at least one content indicator comprises atomic phrases.

BNSDOCID: <WO_____0125947A1_I_>

8. The system of claim 6 or 7, wherein the server system processor further derives the at least one content indicator from each information source by using Relational Methods Language.

9. The system of claim 6, wherein the server system processor recommends the at least one information source based upon the at least one content indicator of the previously-selected information sources of the first user.

10. The system of claim 6, wherein the server system processor recommends the at least one information source based upon the at least one content indicator of the previously-selected information sources of the subset of second users.

11. The system of claim 6, wherein the server system processor recommends at least one of the previously-selected information sources of the first user based upon the at least one content indicator for the previously-selected information sources of the first user.

12. The system of claim 1, wherein the server system processor recommends the at least one information source based upon a frequency of selection for each previously-selected information source of the subset of second users.

13. The system of claim 1, wherein the information sources comprise Internet Web pages.

BNSDOCID: <WO_____0125947A1_I_>

14. The system of claim 1, wherein the first user system processor further communicates a message from the first user to the subset of second users.

15. The system of claim 14, wherein the message is communicated anonymously.

16. The system of claim 14, wherein at least one of the second user system processors communicates at least one response to the first user system processor.

17. The system of claim 16, wherein the at least one response is communicated anonymously.

18. The system of claim 16, wherein the message and the at least one response are communicated via electronic mail over a global computer network.

19. The system of claim 1, wherein the first user system processor and at least one second user system processor communicate, in real-time, two-way messages between the first user and at least one second user from the subset of second users via the communication network.

20. The system of claim 19, wherein the messages are communicated anonymously.

21. The system of claim 19, wherein the messages are communicated via point-to-point instant messaging protocols.

BNSDOCID: <WO_____0125947A1_I_>

22. The system of claim 1, wherein the communication network is selected from the group of communication networks consisting of a global computer network, the Internet, an intranet, a local area network, a wide area network, and a distributed network.

23. A method of dynamically recommending at least one new information source to a first user, the method comprising the steps of

creating a first histogram for the first user, the first histogram comprising

demographic information about the first user; and

historical information identifying previously-selected information sources of the first user;

transforming the first histogram into a first waveform based upon frequencies indicated by the first histogram;

creating a plurality of additional histograms, each additional histogram corresponding to one user of a corresponding plurality of additional users, and each of the additional histograms comprising

demographic information about the corresponding additional user; and

historical information identifying previously-selected information sources of the corresponding additional user;

transforming the plurality of additional histograms into a corresponding plurality of additional waveforms, each additional waveform being based upon frequencies indicated by the respective additional histogram;

comparing the first waveform to the plurality of additional waveforms;

selecting a subset of users from the plurality of additional users, wherein the corresponding additional waveforms of the subset of users indicate an affinity among the first user and the subset of users; and

BNSDOCID: <WO_____0125947A1_I_>

recommending at least one new information source to the first user based upon the previously-selected information sources of the additional users in the subset of users.

24. The method of claim 23, wherein said recommending step further comprises recommending at least one new information source from among the previously-selected information sources of the first user.

25. The method of claim 23, wherein the historical information further comprises at least one content indicator for each previously-selected information source, and wherein the recommending step further comprises recommending the at least one new information source based upon the at least one content indicator for the previously-selected information sources of the first user.

26. The method of claim 23, wherein the historical information further comprises at least one content indicator for each previously-selected information source, and wherein the recommending step further comprises recommending the at least one new information source based upon the at least one content indicator for the previously-selected information sources of the subset of users.

27. The method of claim 23, wherein the historical information further comprises at least one content indicator for each previously-selected information source, and wherein the recommending step further comprises recommending at least one of the previously-selected information sources of the first user based upon the at least one content indicator for the previously-selected information sources of the first user.

BNSDOCID: <WO_____0125947A1_I_>

28. The method of claim 23, wherein the recommending step further comprises recommending the at least one new information source based upon a frequency of selection for each previously-selected information source of the subset of users.

29. The method of claim 23, further comprising the steps of communicating a message from the first user to the subset of users.

30. The method of claim 23, further comprising the step of communicating, in real-time, two-way messages between the first user and at least one user in the subset of users.

31. A method of dynamically recommending at least one new information source to a first user, the method comprising the steps of

creating a first histogram for the first user, the first histogram comprising

demographic information about the first user; and

historical information identifying previously-selected information sources of the first user;

transforming the first histogram into a first waveform based upon frequencies indicated by the first histogram;

creating a plurality of additional histograms, each additional histogram corresponding to one user of a corresponding plurality of additional users, and each of the additional histograms comprising

demographic information about the corresponding additional user; and

historical information identifying previously-selected information sources of the corresponding additional user;

BNSDOCID: <WO_____0125947A1_I_>

transforming the plurality of additional histograms into a corresponding plurality of additional waveforms, each additional waveform being based upon frequencies indicated by the respective additional histogram;

comparing the first waveform to the plurality of additional waveforms;

selecting a subset of users from the plurality of additional users, wherein the corresponding additional waveforms of the subset of users indicate an affinity among the first user and the subset of users; and

communicating a message from the first user to the subset of users, wherein the message requests recommendations for at least one new information source from the subset of users.

32. The method of claim 29 or 31, wherein the message is communicated anonymously.

33. The method of claim 29 or 31, further comprising the step of communicating at least one response to the first user from at least one user in the subset of users.

34. The method of claim 33, wherein the at least one response is communicated anonymously.

35. The method of claim 33, wherein the message and the at least one response are communicated via electronic mail over a global computer network.

36. A method of dynamically recommending at least one new information source to a first user, the method comprising the steps of

BNSDOCID: <WO_____0125947A1_I_>

creating a first histogram for the first user, the first histogram comprising

demographic information about the first user; and

historical information identifying previously-selected information sources of the first user;

transforming the first histogram into a first waveform based upon frequencies indicated by the first histogram;

creating a plurality of additional histograms, each additional histogram corresponding to one user of a corresponding plurality of additional users, and each of the additional histograms comprising

demographic information about the corresponding additional user; and

historical information identifying previously-selected information sources of the corresponding additional user;

transforming the plurality of additional histograms into a corresponding plurality of additional waveforms, each additional waveform being based upon frequencies indicated by the respective additional histogram;

comparing the first waveform to the plurality of additional waveforms;

selecting a subset of users from the plurality of additional users, wherein the corresponding additional waveforms of the subset of users indicate an affinity among the first user and the subset of users; and

communicating, in real-time, two-way messages between the first user and at least one user in the subset of users, wherein the messages comprise requests for at least one new information source from the subset of users.

37. The method of claim 30 or 36, wherein the messages are communicated anonymously.

BNSDOCID: <WO_____0125947A1_I_>

38. The method of claim 30 or 36, wherein the messages are communicated via point-to-point instant messaging protocols over a global system network.

39. The method of claim 23, 31, or 36, wherein the historical information further comprises a specific time of selection for each previously-selected information source.

40. The method of claim 23, 31, or 36, wherein the historical information further comprises a duration of selection for each previously-selected information source.

41. The method of claim 23, 31, or 36, wherein the historical information further comprises a frequency of selection for each previously-selected information source.

42. The method of claim 23, 31, or 36, wherein the historical information further comprises a sequence of selection for each previously-selected information source.

43. The method of claim 23, 31, or 36, wherein the historical information further comprises at least one content indicator for each previously-selected information source.

44. The method of claim 43, wherein the at least one content indicator comprises atomic phrases.

45. The method of claim 43, further comprising deriving the at least one content indicator from each information source by using Relational Methods Language.

BNSDOCID: <WO_____0125947A1_I_>

46. The method of claim 23, 31, or 36, wherein the information sources are selected from the group consisting of Internet Web pages, Internet protocol packets, television vertical blanking intervals, and television horizontal blanking intervals.

47. A method in a computer system for displaying an Internet messaging interface, the method comprising the steps of

displaying a window comprising a graphical user interface for an Internet search engine; and

displaying within the Internet search engine window a communication interface window for communicating messages between a plurality of users of the Internet search engine.

48. The method of claim 47, wherein the communication interface displays messages anonymously.

49. The method of claim 47, wherein the communication interface comprises an electronic mail interface.

50. The method of claim 47, wherein the communication interface provides for real-time, two-way messages between the plurality of users.

51. The method of claim 50, wherein the communication interface comprises point-to-point instant messaging protocols over the Internet.

BNSDOCID: <WO_____0125947A1_I_>

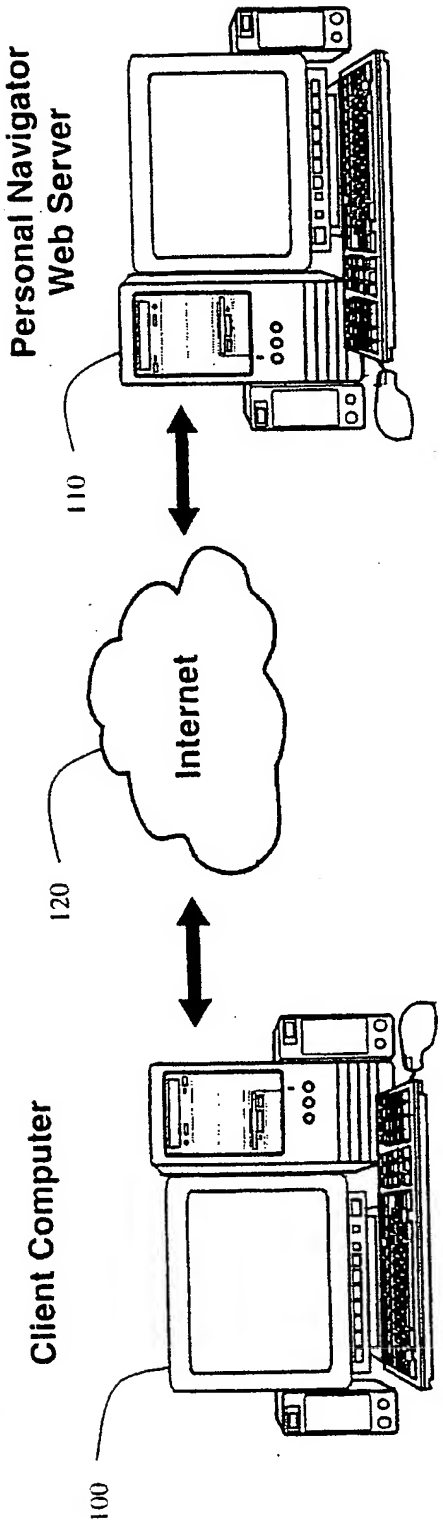


Fig. 1

BNSDOCID: <WO_____0125947A1_1_>

Fig. 2A

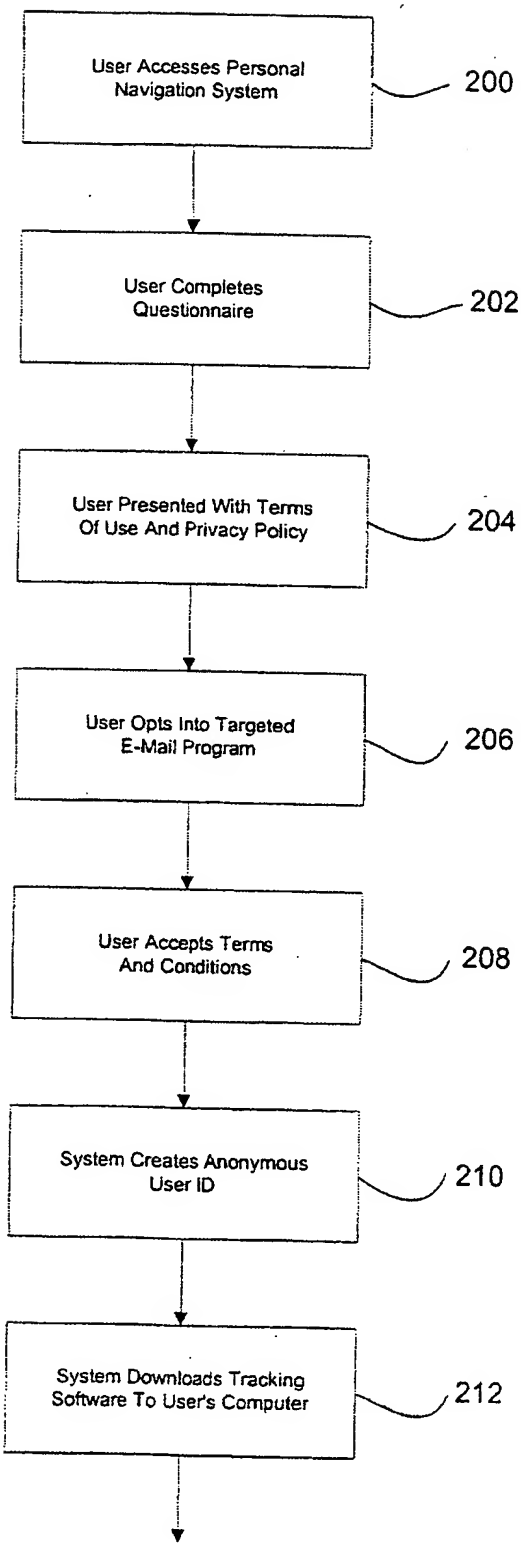

User Accesses Personal Navigation System
200
User Completes Questionnaire
202
User Presented With Terms Of Use And Privacy Policy
204
User Opts Into Targeted E-Mail Program
206
User Accepts Terms And Conditions
208
System Creates Anonymous User ID
210
System Downloads Tracking Software To User's Computer
212

Fig. 2B

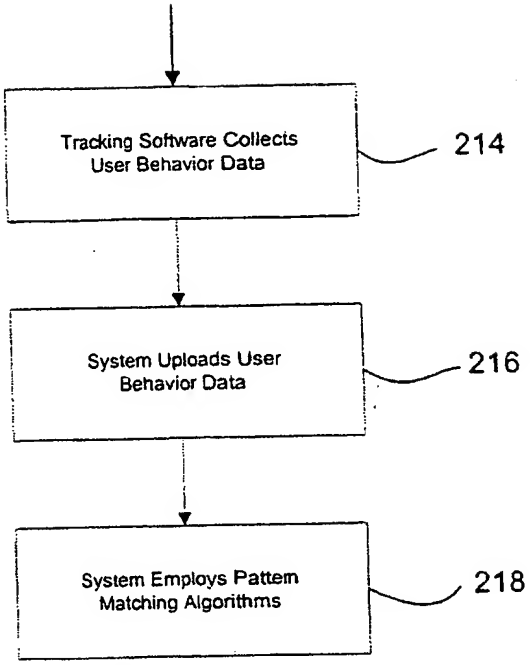

Tracking Software Collects User Behavior Data
214
System Uploads User Behavior Data
216
System Employs Pattern Matching Algorithms
218

Fig. 2C

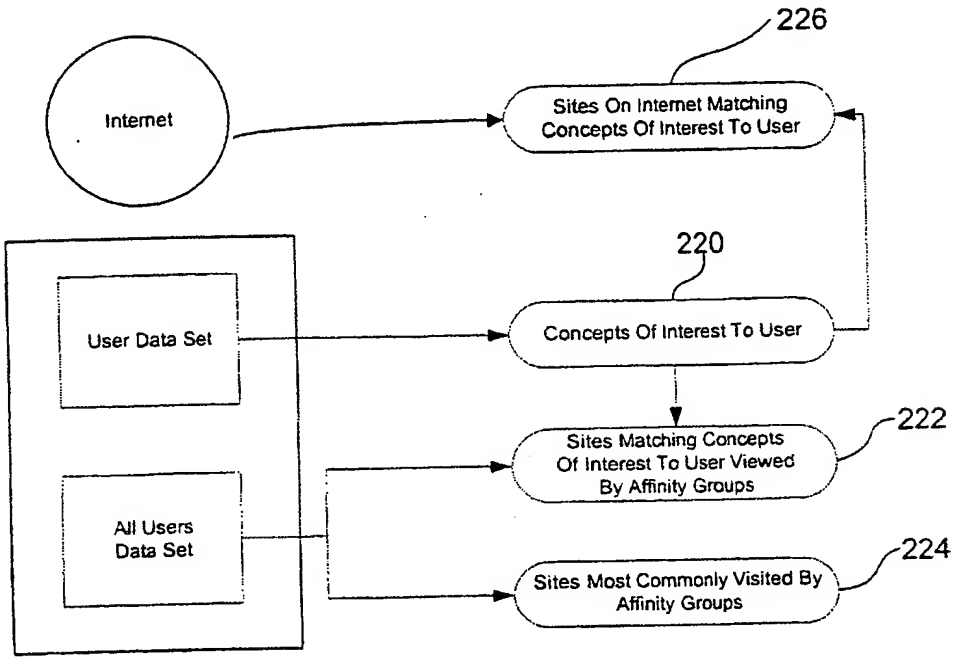
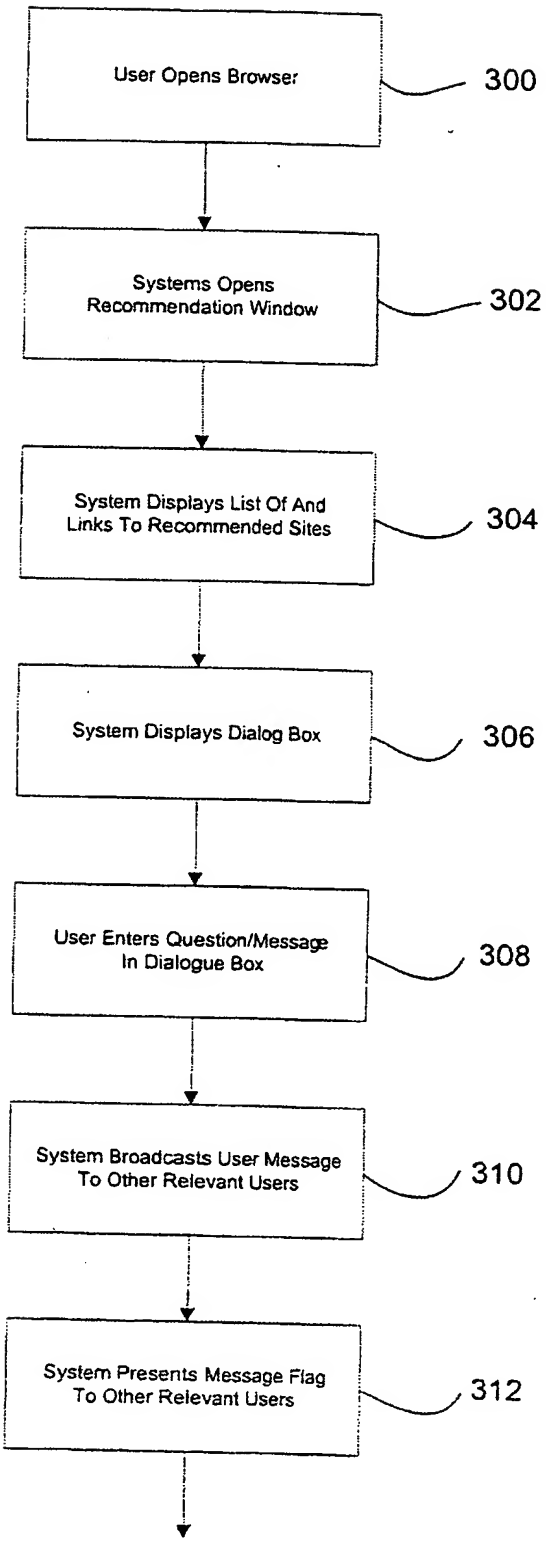

226
Internet
Sites On Internet Matching Concepts Of Interest To User
220
User Data Set
Concepts Of Interest To User
Sites Matching Concepts Of Interest To User Viewed By Affinity Groups
222
All Users Data Set
Sites Most Commonly Visited By Affinity Groups
224

Fig. 3A



#NSDOCID: <WO_____0125947A1_I_>

Fig. 3B

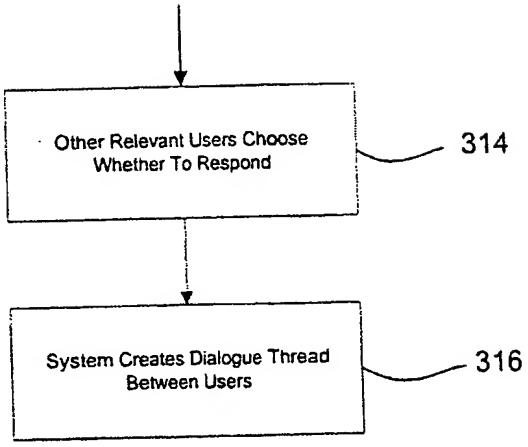
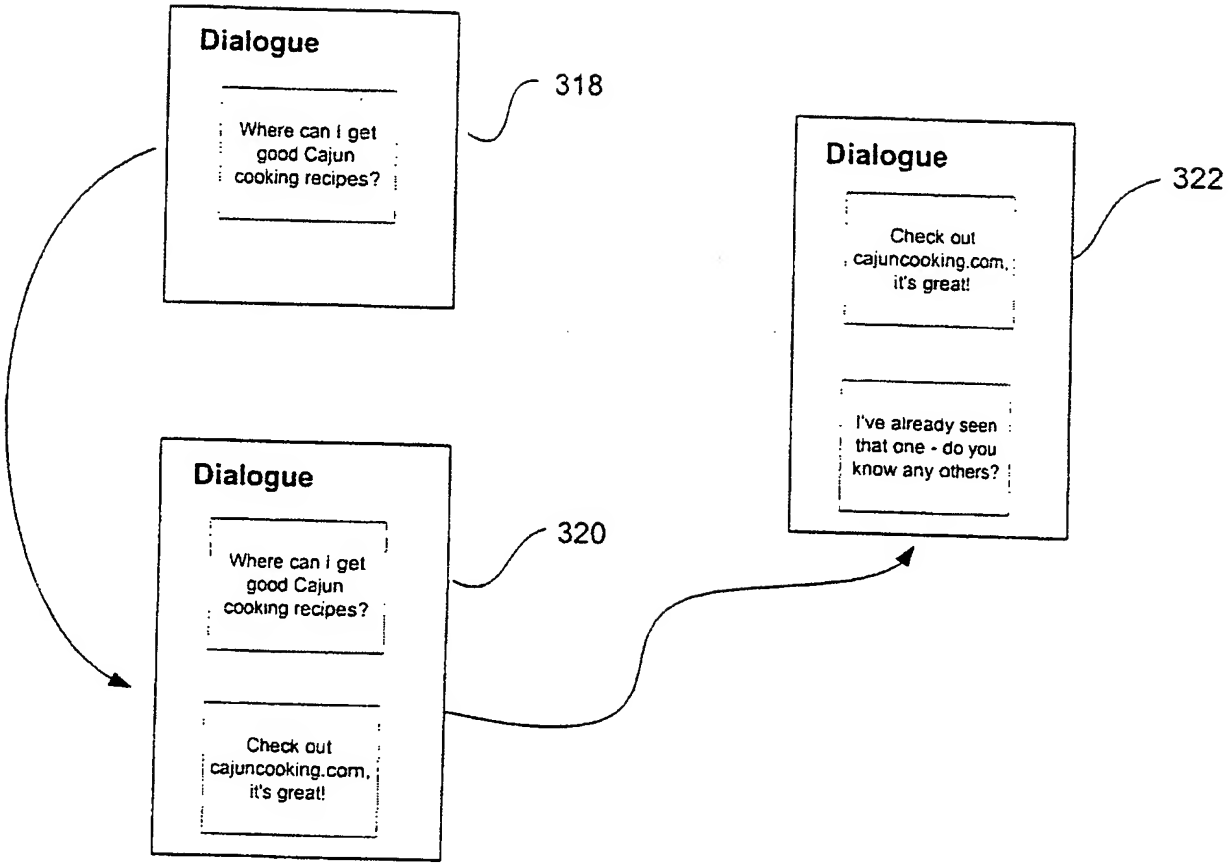
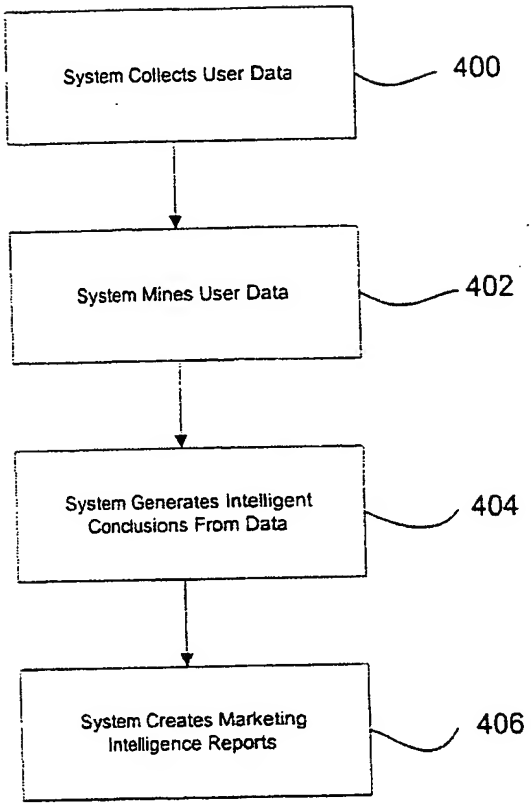

Other Relevant Users Choose Whether To Respond
314
System Creates Dialogue Thread Between Users
316

Fig. 3C



BNSDOCID: <WO_____0125947A1_I_>

Fig. 4


System Collects User Data
400
System Mines User Data
402
System Generates Intelligent Conclusions From Data
404
System Creates Marketing Intelligence Reports
406

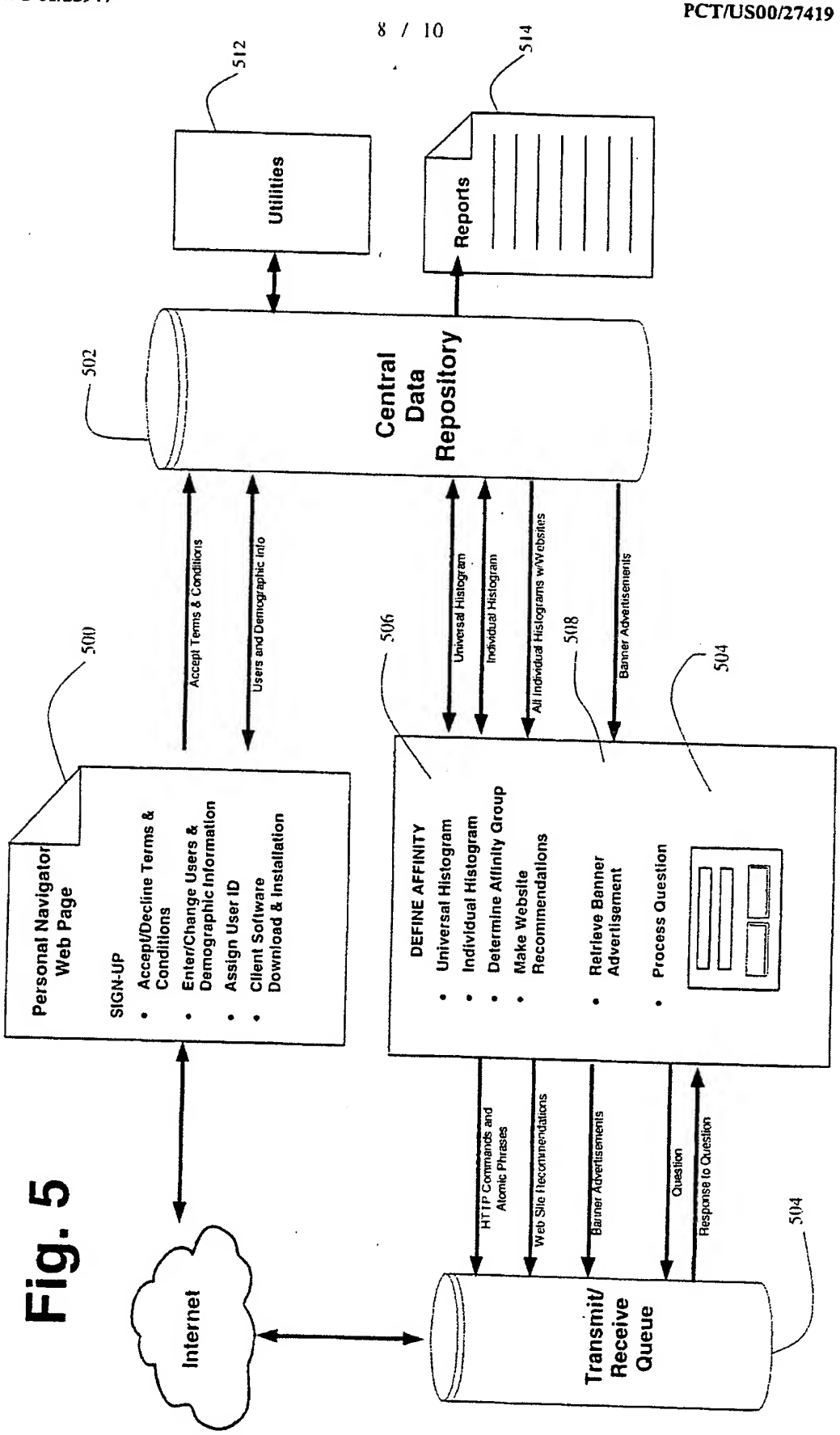


Fig. 5

BNSDOCID: <WO_____0125947A1_I_>

Fig. 7

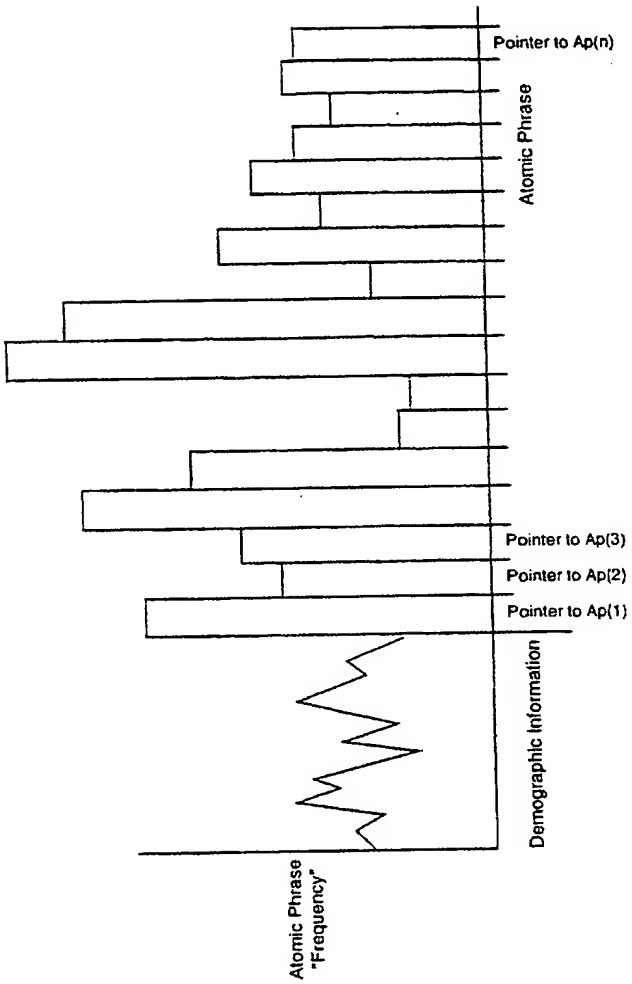
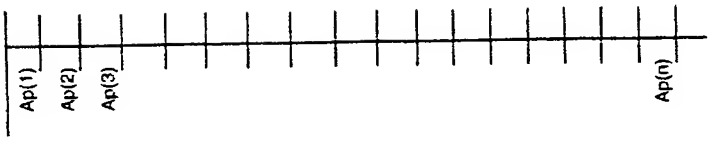

Pointer to Ap(n)
Atomic Phrase
Pointer to Ap(3)
Pointer to Ap(2)
Pointer to Ap(1)
Demographic Information
Atomic Phrase "Frequency"

Fig. 6


Ap(1)
Ap(2)
Ap(3)
Ap(n)

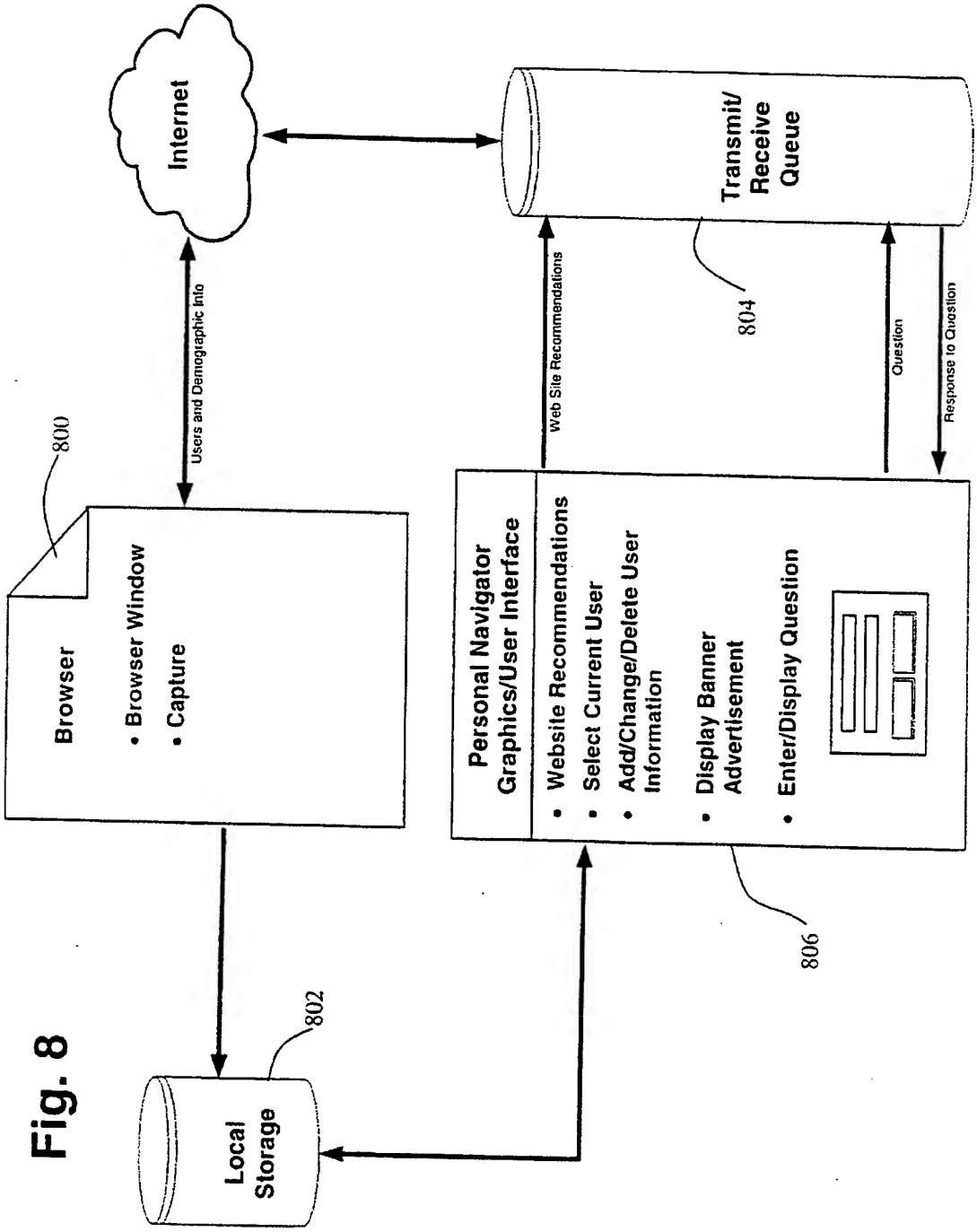


Fig. 8

BNSDOCID: <WO_____0125947A1_I_>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/27419

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G06F 15/16,
US CL : 709/205, 217

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 709/205, 217

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

IEEE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EAST, STN

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,931,907 A (DAVIES et al.) 03 August 1999, col. 3-9	1-51
Y, P	US 6,009,410 A (LEMOLE et al.) 28 December 1999, col. 2-7	1-51
Y	US 5,933,811 A (ANGLES et al.) 03 August 1999, 6-22	1-51

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 02 DECEMBER 2000	Date of mailing of the international search report 09 JAN 2001
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer SALEH NAJJAR James R. Matthews Telephone No. (703) 308-7613

Form PCT/ISA/210 (second sheet) (July 1998)*

BNSDOCID: <WO_____0125947A1_I_>